

Predicted Effector Molecules in the Salivary Secretome of the Pea Aphid (*Acyrtosiphon pisum*): A Dual Transcriptomic/Proteomic Approach

James C. Carolan,^{*,†,‡} Doina Caragea,[§] Karen T. Reardon,[†] Navdeep S. Mutti,^{||,⊥} Neal Dittmer,[⊥] Kirk Pappan,^{⊥,¶} Feng Cui,^{⊥,||} Marisol Castaneto,[▲] Julie Poulain,[▲] Carole Dossat,[▲] Denis Tagu,[○] John C. Reese,[●] Gerald R. Reeck,[⊥] Thomas L. Wilkinson,[†] and Owain R. Edwards^{||,□}

[†]UCD School of Biology and Environmental Science, University College Dublin, Dublin 4, Ireland

[§]Department of Computer Science and Information, and Bioinformatics Center, Kansas State University, Manhattan, Kansas 66506, United States

^{||}School of Life Sciences, Arizona State University, Tempe, Arizona 85287, United States

[⊥]Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506, United States

[¶]Cooperative Research Centre for National Plant Biosecurity, Canberra, ACT 6201 Australia

[▲]Genoscope, CEA, DSV, Institut de Génomique, 2 rue Gaston Crémieux 91000 Evry, France

[○]INRA Rennes, UMR1099 BiO3P, Domaine de la Motte, F-35653 Le Rheu, France

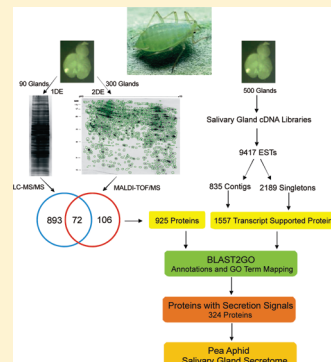
[●]Department of Entomology, Kansas State University, Manhattan, Kansas 66506, United States

[□]CSIRO Entomology, Centre for Environment and Life Sciences, Floreat, WA 6014 Australia

S Supporting Information

ABSTRACT: The relationship between aphids and their host plants is thought to be functionally analogous to plant–pathogen interactions. Although virulence effector proteins that mediate plant defenses are well-characterized for pathogens such as bacteria, oomycetes, and nematodes, equivalent molecules in aphids and other phloem-feeders are poorly understood. A dual transcriptomic–proteomic approach was adopted to generate a catalog of candidate effector proteins from the salivary glands of the pea aphid, *Acyrtosiphon pisum*. Of the 1557 transcript supported and 925 mass spectrometry identified proteins, over 300 proteins were identified with secretion signals, including proteins that had previously been identified directly from the secreted saliva. Almost half of the identified proteins have no homologue outside aphids and are of unknown function. Many of the genes encoding the putative effector proteins appear to be evolving at a faster rate than homologues in other insects, and there is strong evidence that genes with multiple copies in the genome are under positive selection. Many of the candidate aphid effector proteins were previously characterized in typical phytopathogenic organisms (e.g., nematodes and fungi) and our results highlight remarkable similarities in the saliva from plant-feeding nematodes and aphids that may indicate the evolution of common solutions to the plant–parasitic lifestyle.

KEYWORDS: *Acyrtosiphon pisum*, aphids, salivary glands, saliva, effectors



INTRODUCTION

As primary producers, plants are consumed by a wide variety of generalist and specialist insect herbivores. The interplay between plants and their consumers is highly complex and dynamic involving interactions between the insect and a plethora of constitutive and inducible defensive mechanisms that have evolved to afford the plant a degree of “dynamic immunity” (reviewed in refs 1–4). The plant response to herbivory often depends on the mechanism of feeding employed by the consumer. For example, leaf chewing insects such as lepidopteran larvae consume large pieces of tissue consisting of many cells, and

produce a plant response that is very different from the response to insects such as aphids, whiteflies and mealybugs that penetrate between plant cells to feed from individual phloem cells within the sieve elements. There is a growing body of evidence that plant defense against sap-feeding insects such as aphids follows the plant-pathogen model as presented by Jones and Dangl.⁵ For instance, aphid-induced gene expression in the plant is more similar to a typical pathogen response than to that caused by a chewing insect,⁶ reviewed in ref 7. In addition the only cloned

Received: August 28, 2010

Published: January 12, 2011

aphid resistance genes, the *Mi* gene in tomato⁸ and the *Vat* gene in melon,⁹ are members of the CC-NBS-LRR family of resistance proteins (*R* proteins) that confer resistance to both prokaryotic and eukaryotic pathogens.¹⁰ Several other aphid resistance genes have been mapped to clusters of *R* genes.^{11,12} Some aphid species have also adapted to overcome resistance deployed in commercial crops, forming virulent biotypes specific to particular *R* genes following the “gene for gene” pathogen avirulence model.¹³

Under the plant–pathogen model, compatible hosts are manipulated by the secretion of virulence effector molecules that suppress or otherwise interfere with the plant’s innate immunity defenses (reviewed in refs 5, 14). Many pathogens have specific secretion systems that enable the delivery of these effectors, such as the type III secretion system in bacteria^{15–17} or the RXLR system in oomycetes.^{18–20} By the nature of their feeding, aphids are thought to deliver effectors in their saliva,^{21–23} as is the case for plant pathogenic nematodes.^{24–26} However, little is known about the bioactive components of aphid saliva. Aphids initially secrete a gelling saliva that hardens and forms a protective sheath around the mouthparts, followed by watery saliva which is secreted into the phloem immediately before and during feeding but also into cells as the tip of the mouthparts progress toward the phloem.^{21,22}

A number of strategies have been employed in an attempt to catalogue and characterize the salivary proteins of various aphid species and include mass spectrometry of collected saliva,^{27,28} RNA interference,²⁹ salivary gland EST analysis,³⁰ *in planta* analysis of phloem occlusion mechanisms,²³ plant differential gene expression studies³¹ in response to aphid salivary concentrates and *in planta* expression of candidate salivary effector proteins.³² The availability of a genome sequence for the pea aphid, *Acyrtosiphon pisum*, has made this species a particularly good candidate for saliva studies. Using antibody-based approaches, Mutti et al.³³ confirmed that a protein called C002, initially identified by prospecting salivary gland ESTs and genome analysis, was detected in the plants upon which aphids had fed, confirming that it is a component of the secreted saliva. More recently, mass spectrometry (LC–MS/MS) resulted in the identification of nine secreted salivary proteins, four of which exhibited orthology to proteins in existing databases:²⁸ a GMC oxidoreductase, an SMP-30-like protein (regucalcin), an angiotensin converting enzyme (Ance)-like protein (carboxyl dipeptidase) and an M1 metalloprotease.

Here we characterize the salivary secretome of the pea aphid (*Acyrtosiphon pisum*) through a parallel analysis of proteins and transcripts in the salivary glands to gain further insights into the origin and composition of aphid saliva, particularly the identification of potential effector molecules. Aphids are economically important pests of temperate agriculture that inflict damage and loss of yield through the depletion of photoassimilates and the transmission of plant viruses. Consequently, a deeper understanding of the role of saliva in the aphid–plant interaction has significant applied potential. To characterize the pea aphid salivary secretome, we adopted three complementary approaches. First, we analyzed ESTs from two pea aphid salivary gland cDNA libraries to identify transcripts that were significantly over-represented in the salivary glands. Second, we performed mass spectrometry on proteins isolated from both 1D and 2D separations of salivary gland homogenates. Reference genes corresponding to the two data sets that were identified from aphid genomic gene consensus sets (www.aphidbase.com) were analyzed for the presence of a secretion signal peptide. Figure 1

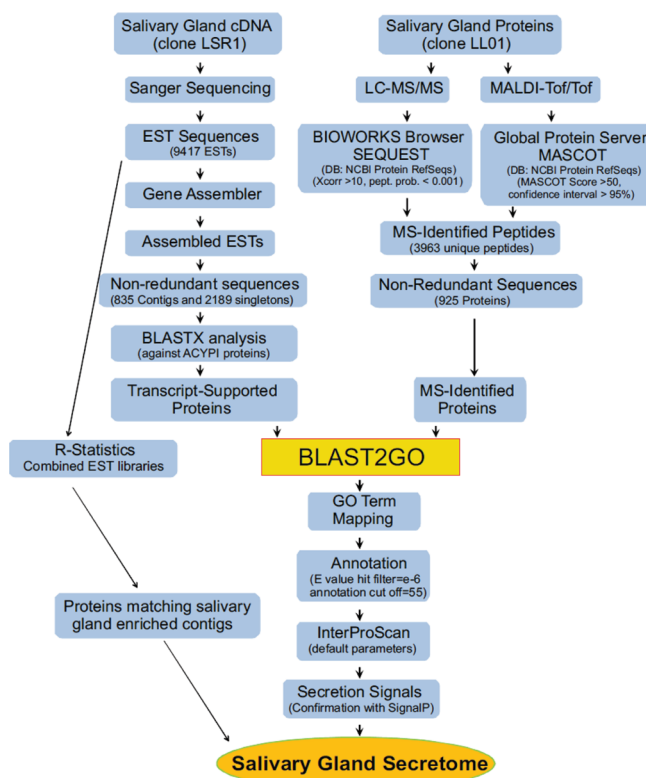


Figure 1. Schematic representation of the experimental pipeline to characterize the pea aphid salivary gland secretome detailing the experimental steps and statistical parameters. Blast2Go was utilized to annotate non redundant EST contigs and singletons and MS-identified proteins and assign GO terms to the positive BLAST matched proteins. The constituents of the salivary gland secretome were selected based on presence of signal peptide sequence, support by EST contigs with a high R-Stat value and previous demonstration of association with aphid saliva.

outlines an overview of the experimental steps, statistical parameters and data treatment involved in the generation of the pea aphid salivary gland secretome. Lastly, we conducted a phylogenetic analysis on the most well-supported secreted salivary protein candidates to identify putative paralogs within the pea aphid genome and to determine their homology to genes in other insects.

MATERIALS AND METHODS

Aphids, Salivary Glands, RNA and Protein Collection

Acyrtosiphon pisum clone LL01³⁴ was maintained on *Vicia faba* (cv. The Sutton) at University College Dublin, Ireland and *A. pisum* clone LSR1 was maintained on *Vicia faba* (cv. Broad Windsor) at Kansas State University, Manhattan, KS. Salivary glands were obtained from adult apterae of each colony following a modified protocol of.²⁹ For RNA extraction, salivary glands (from aphid clone LSR1) were dissected in PBS supplemented with protease inhibitor cocktail (Roche) and total RNA was isolated using the Micro RNA isolation kit (Stratagene) following the manufacturer’s instructions. For protein analysis (using aphid clone LL01), salivary glands were dissected out in lysis buffer (LB; 9.5 M urea; 2% CHAPS; 8% Pharmalyte pH 3–10 and 1% DTT). Salivary glands were homogenized in LB prior to sonication (Vibra Cell Ultra Sonic Processor, Sonics, U.K.) at 4 °C by 5 × 10-s bursts. Samples were subjected to 2D-clean up

(2D-clean up kit; GE Healthcare) following the manufacturer's instructions and the final pellet was solubilized in 100 μ L LB. 90 and 300 pairs of salivary glands were pooled into 100 μ L LB for the SDS PAGE (1DE) and 2-dimensional electrophoresis (2DE) fractionation steps, respectively.

Phagmid cDNA Library Construction, EST Sequencing and Analysis

Two phagmid salivary gland cDNA libraries were constructed (see Supplemental File 1, Supporting Information for further details). The first was constructed from 250 dissected salivary glands using the SMART cDNA library construction kit (Clontech) as described previously.^{29,33} Sequencing of this cDNA library was conducted at Genoscope (Paris, France), generating 4517 ESTs (accession numbers DV747494–DV752010). Sequencing of the second library, again generated from 250 salivary glands, was conducted at the Washington University Sequencing Facility (St. Louis, MO), generating 4900 ESTs (accession numbers HS092103–HS097002). Thus, a total of 9417 ESTs were analyzed.

EGassembler³⁵ was used to assemble ESTs into contiguous sequences (contigs) and unassembled reads (singletons). Details of the steps and parameters used in EGassembler are provided in Supplemental File 1 (Supporting Information). To obtain a more reliable and comprehensive EST-contig set, EGassembler was also used to assemble all pea aphid ESTs available at NCBI into a combined EST library. This approach generated a total of 99 776 ESTs comprising 963 from an antennae library; 1109 from a gut library; 12 972 from three head libraries; 5443 from an embryo library; 16 759 from 4 non-normalized whole-body libraries; 53 113 from a normalized whole-body library; and 9417 from the two salivary-gland libraries reported in this study.

LC–MS/MS and MALDI-TOF/MS

Seventy-five micrograms and 400 μ g of salivary gland homogenates were used for the 1DE gel/LC–MS/MS and 2D gel/MALDI-TOF/MS respectively. Details of in gel trypsin digestion, mass spectrometer settings, data acquisition and search parameters are provided in Supplemental File 1 (Supporting Information). Proteins identified from the 2DE/MALDI-TOF/MS data set were compared with the 1DE fractionated LC–MS/MS identified proteins to determine overlap between fractionation methods. Redundant protein hits between the two data sets were removed from the 2DE protein list prior to the combination of all proteins into a single data set (hereafter mass spectrometry (MS)-identified proteins). The genome derived databases used for peptide/protein searches were produced and made available by the Human Genome Sequencing Centre at Baylor College of Medicine (www.hgsc.bcm.tmc.edu/) and the International Aphid Genomics Consortium (IAGC).

BLAST Searches, Functional Annotation and Secretion Signal Prediction

Preliminary annotation of the identified proteins, EST contigs and singletons was performed using Blast2GO (ref 36, v.2; <http://www.blast2go.org/>) an integrative GO annotation and data mining suite that assigns GO annotation through BLAST³⁷ searches against nucleotide and protein databases. Further details on BLAST2GO analyses and database searches are provided in Supplemental File 1, Supporting Information. Assigned GO terms for the salivary gland ESTs, transcript-supported and the MS-identified proteins were categorized by molecular function (MF), biological process (BP) and cellular component (CC) to

suggest the functional component of the salivary gland proteome and transcriptome. Signal secretion peptides were identified from the results arising from the InterProScan searches and all proteins with identified signal peptides were analyzed further using SignalP v. 3.0 (<http://www.cbs.dtu.dk/services/SignalP>) to predict a potential secretion signal peptide sequence and to determine cleavage site position. Transcript-supported and all MS-identified proteins that comprised a secretion signal were categorized by MF, BP and CC to determine the functional component of the potentially secreted protein complement of the pea aphid.

Salivary Gland Enriched Transcripts

Using the whole organism ESTs, R-statistic values were calculated to identify EST contigs abundant in the salivary gland following the methods of Stekel et al.³⁸ The R-statistic examines the null hypothesis that the frequency of a transcript (i.e., contig) is the same in all other non-normalized libraries. This null hypothesis is rejected at an R-statistic value of 7 and EST contigs with R-statistic values greater than 7 are considered to occur at elevated levels in salivary-glands to a statistically significant extent. Contigs determined to be statistically over-represented in the salivary gland were BLASTx searched against the ACYPI-protein consensus set, permitting matching of these contigs to the transcript-supported and MS-identified proteins.

Phylogenetic Analysis of Candidate Secreted Salivary Proteins

Predicted salivary proteins were considered well-supported if they were identified as candidates by at least two independent approaches: salivary gland EST analysis, salivary gland proteome analyses, and/or MS identification in secreted saliva.²⁸ The best-supported of the predicted secreted salivary proteins were subjected to PhylomeDB analysis (<http://phylomedb.org>) to identify related genes arising by speciation (orthologs) or duplication (paralogs). Putative paralogs were cross-referenced to the EST and proteome data sets (salivary gland and saliva) to determine whether there was any evidence that the paralogs could also be secreted salivary proteins. Each matched ACYPI accession for the non annotatable proteins was searched using PhylomeDB to determine whether these proteins were specific to aphids or conserved across other insect species.

Ortholog and paralog amino acid sequences were collected from PhylomeDB and aligned using ClustalX,³⁹ including one outgroup (from *Daphnia*, *Caenorhabditis elegans*, or *Homo sapiens*). The corresponding alignments of the coding sequences (CDS), obtained from insect genome Web sites or from NCBI (Entrez Gene), were generated using Codon Align 2.0 (<http://sinauer.com/hall/2e/>). Information on rates of evolution and phylogenetic analyses is provided in Supplemental File 1, Supporting Information.

RESULTS

Salivary Gland EST Characteristics

The combined assembly of the individual salivary gland EST libraries (denoted SG-EST1 and SG-EST2, respectively) resulted in 835 contigs and 2189 singletons (unassembled reads). General statistics concerning the raw unassembled salivary gland ESTs and the contigs and singletons for the individual and combined salivary gland EST libraries are provided in Supplemental File 2, Supporting Information. The assembly of the combined EST library resulted in 11 009 contigs and 11 619

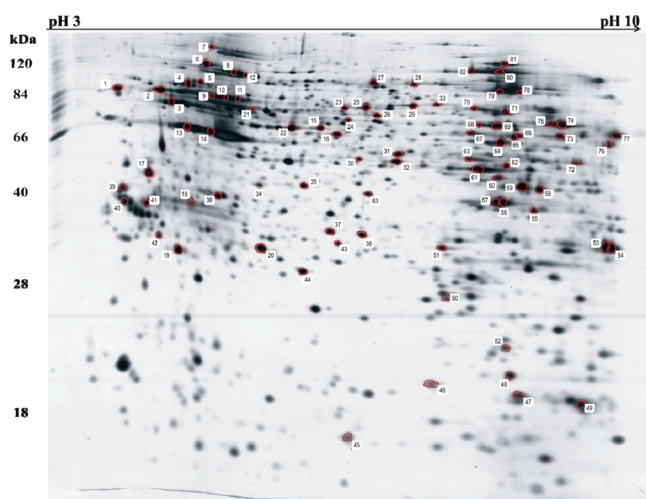


Figure 2. Silver stained 2D gel image of total protein from salivary glands of *Acyrthosiphon pisum*. Proteins were separated on a 3–10 NL IPG strip in the first dimension and on an SDS-PAGE (12%) in the second dimension. Protein identities corresponding to spot numbers are provided in Supplemental File 3, Supporting Information. In total, 151 proteins were identified using MALDI-TOF/MS.

singletons. Out of the 11 009 contigs, 1486 contained at least one salivary gland EST, and out of the 11 619 singletons, 1569 were from the salivary gland EST libraries.

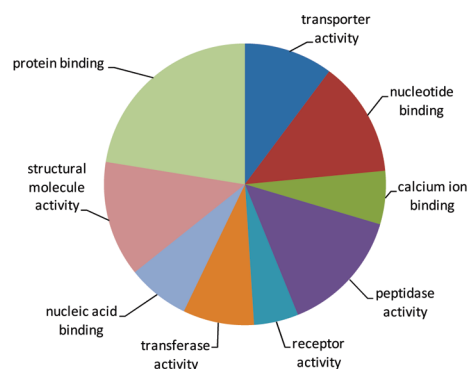
LC–MS/MS and MALDI-TOF/MS of the Pea Aphid Salivary Gland

1DE and 2DE fractionated proteins were subjected to LC–MS/MS and MALDI-TOF/MS respectively and identified after SEQUEST and MASCOT searches of MS/MS spectra against data sets originating from the Pea Aphid Genome Sequencing Project. In total 3661 unique peptides representing 893 non redundant proteins were identified from the 25 gel pieces cut from the preparative 1DE gel (Supplemental File 3, Supporting Information). Of these, 609 proteins were supported by more than one peptide. Over 300 proteins were detectable on the silver stained 2-D gel (Figure 2) with the majority of proteins having molecular masses between 120 and 10 kDa. Of the 194 spots excised from the 2D gel, identities could be assigned to 122 of these after MS/MS searches against the non redundant NCBI database and pea aphid consensus protein databank. A number of peptides from different spots matched to the same protein (indicative of isoforms or differential post translational modification patterns) and a number of spots comprised more than one protein. The number of unique proteins identified was 106 and the full set of protein hits corresponding to the spot numbers outlined in Figure 2 is provided as Supplemental File 3, Supporting Information. 72 proteins were shared between the 1DE LC–MS/MS and 2DE MALDI TOF/MS data sets (including two hits to different splice variants) and the combination of both data sets resulted in 925 MS-identified proteins.

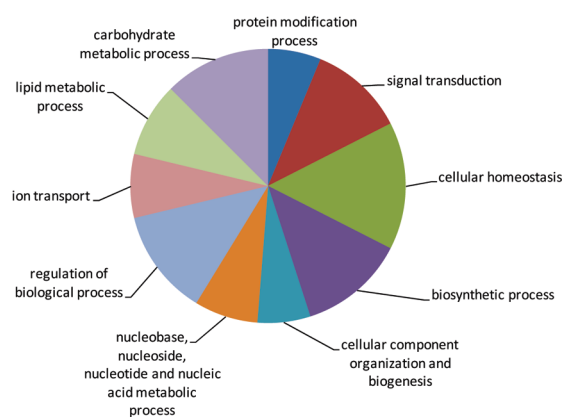
Annotation of Salivary Gland Transcripts and Proteins

Of the 835 contiguous EST sequences 473 and 621 returned significant BLAST hits when searched against the NCBI nr and official aphid ACYPI protein databanks, respectively (Supplemental Files 4 and 5, Supporting Information). 1296 of the 2189 EST singletons returned significant BLAST hits when searched against the official ACYPI gene databank (with 850 significant hits when searched against the NCBI nr database). The vast

MF



BP



CC

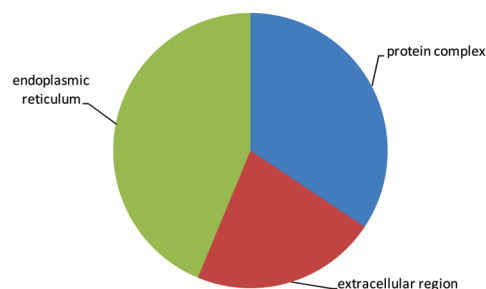


Figure 3. Gene Ontology (GO) terms for each of the three main categories molecular function (MF), biological process (BP), and cellular component (CC) obtained for the pea aphid salivary gland secretome.

majority of contigs and singletons matched to pea aphid protein sequences (Supplemental File 5, Supporting Information). Of the 1917 contigs and singletons with a supported BLAST hit, 1557 had a corresponding aphid protein hit (transcript-supported proteins) and were predicted to have a complete sequence. Of the 925 MS-identified proteins 81 yielded no significant BLAST hits or matched to proteins with unknown function and were deemed nonannotatable (Supplemental File 6, Supporting Information). One-hundred twenty-five of the 621 BLAST matched contigs and 224 of the 851 BLAST matched EST singletons were deemed nonannotatable.

A functional overview of the salivary gland proteome was obtained through the analysis of identified proteins using BLAST2GO. GO term mapping was based on sequence similarity to previously GO mapped sequences available in the UniProt database and by merging GOs identified after the InterProScan searches of all sequences. 2057 and 2831 GO terms were assigned to

Table 1. Forty-two Proteins of the Predicted Pea Aphid Salivary Gland Secretome Supported by Abundant Salivary Gland EST Contigs^a

protein hit	sequence description	contig ID	R-stat Value	BLAST e-value	BLAST bit score	MS-identified	protein score	no. of peptides
ACYPI008617-PA	hypothetical protein	3433	479.1	5.00×10^{-113}	404	YES	86.31	9
		3222	50.8	4.00×10^{-113}	404			
ACYPI009881-PA	hypothetical protein	3368	188.8	4.00×10^{-105}	379	YES	310.3	31
ACYPI000472-PA	hypothetical protein	3169	193.0	5.00×10^{-107}	385	YES	40.23	4
ACYPI008224-PA	hypothetical protein	3447	158.2	3×10^{-84}	308	YES	30.3	3
ACYPI000490-PA	hypothetical protein	3302	121.3	6.00×10^{-117}	394	YES	86.29	9
ACYPI006346-PA	hypothetical protein	3466	120.5	2×10^{-68}	256	YES	30.28	3
ACYPI001271-PB	hypothetical protein	3252	77.2	2.00×10^{-76}	282	YES	10.21	1
		2958	28.6	2.00×10^{-102}	369			
ACYPI49603-PA	hypothetical protein	3197	73.0	8.00×10^{-49}	194	NO		
ACYPI002439-PA	glutathione peroxidase	3455	70.8	1.00×10^{-135}	479	YES	10.25	1
ACYPI39568-PA	hypothetical protein	3291	67.1	0	791	NO		
ACYPI007406-PA	hypothetical protein	3470	51.8	5.00×10^{-113}	403	YES	20.32	2
		1585	30.8	4.00×10^{-104}	374			
		3523	25.4	1.00×10^{-113}	405			
ACYPI45001-PA	hypothetical protein	3400	47.6	1.00×10^{-80}	296	NO		
ACYPI000852-PA	hypothetical protein	3298	40.2	3.00×10^{-72}	268	NO		
ACYPI000223-PA	hypothetical protein	3253	36.0	1.00×10^{-103}	372	YES	20.14	2
ACYPI000288-PA	glucose dehydrogenase	3485	34.1	0	1328	YES	118.27	12
ACYPI55147-PA	hypothetical protein	3200	29.6	0	1481	YES	20.22	2
ACYPI000733-PA	angiotensin converting enzyme	10169	28.6	2.00×10^{-102}	368	YES	30.32	3
ACYPI002476-PA	inositol monophosphatase family domain containing protein	3438	28.6	3.00×10^{-81}	298	YES	10.2	1
ACYPI001541-PA	hypothetical protein	3409	24.3	5.00×10^{-73}	265	NO		
ACYPI001606-PA	hypothetical protein	3328	19.0	3.00×10^{-135}	478	YES	10.11	1
ACYPI000986-PA	glucose dehydrogenase	9619	18.0	0	1023	YES	250.28	25
ACYPI001719-PA	hypothetical protein	3467	15.9	3.00×10^{-91}	332	YES	114.35	12
		3388	7.8	1.00×10^{-131}	465			
ACYPI003917-PA	SCP_GAPR-1_like family, cg16995	3157	14.8	8.00×10^{-54}	206	YES	70.32	7
ACYPI003601-PA	conserved hypothetical protein, similar to CG4090	3198	14.8	3.00×10^{-142}	503	YES	30.29	3
ACYPI008667-PA	hypothetical protein	3208	14.8	5.00×10^{-57}	217	NO		
ACYPI002172-PA	hypothetical protein	2964	12.4	3×10^{-18}	80.1	YES	20.16	2
		3201	14.8	2.00×10^{-42}	170	NO		
ACYPI002891-PA	cadherin	3396	14.8	5.00×10^{-65}	243	NO		
ACYPI001099-PA	hypothetical protein	3305	7.4	3.00×10^{-93}	338	NO		
		3361	14.8	5.00×10^{-81}	181	NO		
ACYPI56502-PA	hypothetical protein	3378	10.6	7.00×10^{-78}	287	NO		
ACYPI000558-PA	hypothetical protein, partial	3175	10.8	4.00×10^{-83}	305	YES	10.22	1
ACYPI009919-PA	hypothetical protein	3308	10.6	4.00×10^{-59}	224	NO		
ACYPI001843-PA	hypothetical protein	3175	10.8	1.00×10^{-106}	384	NO		
ACYPI005818-PB	hypothetical protein	3532	10.6	3.00×10^{-53}	204	NO		
ACYPI004198-PA	lipophorin precursor	3287	9.5	4.00×10^{-104}	374	YES	20.28	2
ACYPI008001-PA	endopeptidase inhibitor-like	3357	8.5	8.00×10^{-92}	333	YES	20.16	2
ACYPI003695-PA	hypothetical protein	3178	8.5	4.00×10^{-65}	243	NO		
ACYPI001887-PA	hypothetical protein	3213	8.5	6.00×10^{-99}	357	YES	10.24	1
ACYPI55148-PA	hypothetical protein	3181	8.5	0	1669	YES	10.19	1
ACYPI43360-PA	hypothetical protein	7579	8.5	9.00×10^{-73}	274	NO		
ACYPI007553-PB	hypothetical protein	3269	8.5	1.00×10^{-88}	315	NO		
ACYPI001152-PA	hypothetical protein	1134	7.7	1.00×10^{-92}	337	NO		
ACYPI38795-PA	hypothetical protein	942	7.2	2.00×10^{-32}	141	NO		

^a Contigs were BLASTx searched against the official protein consensus set to identify the corresponding protein. Of the 121 significant R-statistics supported EST contigs 50 matched to proteins with signal peptide sequences, 25 of which were identified directly in salivary glands by mass spectrometry; 32 proteins supported by salivary gland abundant EST contigs are non-annotatable.

Table 2. Proteins from the Pea Aphid Salivary Gland Secretome with Previously Observed Association to Saliva or with Some Similarity to Effectors Secreted by Phytopathogenic Organisms

protein	ACYPI	MS-identified?	transcript-supported?	comments
M1 zinc metalloprotease	ACYPI009427	N	Y	M1Metalloproteases have been demonstrated in the saliva using MS. ²⁸
	ACYPI002258	Y	Y	Numerous metalloproteases have been reported in the secretome of the plant parasitic nematode <i>Meloidogyne incognita</i> . ²⁴
Protein of unknown function C002	ACYPI008617	Y	Y	C002: Essential for successful feeding of aphids from plants by RNAi. ²⁹ Unique to aphids.
Glucose dehydrogenase	ACYPI000288	Y	Y	GMC oxidoreductases have been demonstrated by mass spectrometry in aphid saliva. ^{27,28}
Disulfide isomerase	ACYPI009755	Y	Y	Abundantly represented in the secretome of <i>Meloidogyne incognita</i> characterized by MS. ²⁴
	ACYPI005594	Y	Y	
	ACYPI008926	Y	Y	
	ACYPI000119	Y	y	
SCP GAPR-1	ACYPI003917	Y	N	Member of the SCP protein family which includes plant pathogenesis-related protein 1 (PR-1).
Calreticulin	ACYPI002622	Y	N	A common nematode excretory-secretory protein. ⁷⁶
	ACYPI007677	Y	N	
Protein of unknown function	ACYPI009881	Y	Y	Putative sheath protein: Unique to aphids. Major component of aphid saliva. ²⁸
Angiotensin converting enzyme-like	ACYPI000733	Y	Y	Demonstrated by mass spectrometry in aphid saliva. ²⁸
				A dipeptidyl carboxypeptidase which cleaves short peptides.
				Potential effector function as a processor of signaling peptides.
Protein of unknown function	ACYPI008224	Y	Y	Unique to aphids. Demonstrated in aphid saliva using mass spectrometry. ²⁸
Protein of unknown function	ACYPI006346	Y	Y	Unique to aphids. Demonstrated in aphid saliva using mass spectrometry. ²⁸
ARMET	ACYPI008001	Y	Y	A common nematode excretory-secretory protein. ^{24,77}
Glutathione peroxidase	ACYPI002439	Y	Y	A commonly reported protein in insect salivary glands ⁷⁸ and nematode secretomes. ^{24,76} Involved in detoxification.
Trehalase	ACYPI002298	Y	N	Trehalase is induced in plants during infection by microorganisms. ⁷⁹
CLIP-domain serine protease	ACYPI008370	Y	N	A common nematode excretory/secretory product involved in inhibiting phenoloxidase-based innate defenses.

the MS-identified sequences and transcript-supported proteins respectively. Of the 1557 transcript-supported proteins and 925 MS-identified proteins at least one GO term could be assigned to 922 and 731 respectively (Supplemental Files 5 and 6, Supporting Information). The functional component of MS-identified and transcript-supported proteins was determined by categorizing the GO terms for molecular function (MF), biological process (BP) and cellular component (CC; Supplemental File 7, Supporting Information).

Salivary Gland Enriched Transcripts

Transcripts enriched in salivary-gland cDNA libraries compared to (non-normalized) whole-body cDNA libraries were determined using the R-statistic method of Stekel et al. (2000). In total, 121 contiguous ESTs from pea-aphid whole organism libraries, with positive BLAST matches to genomic loci had salivary-gland R-values of 7 or greater (Supplemental File 8, Supporting Information). Approximately half (59) of the salivary-gland-enriched EST-contigs have strong matches to ESTs from a salivary-gland library of the green peach aphid *Myzus persicae*³⁰ (Supplemental File 8, Supporting Information).

Signal Peptide Prediction and GO Term Mapping of the Salivary Gland Secretome

InterProScans of the MS-identified and transcript-supported protein data sets identified 156 and 273 proteins with predicted secretion signals respectively, representing 373 unique directly identified or transcript-supported proteins, 56 of which were

present in both the MS-identified and transcript-supported protein data sets. Of these 373 proteins, SignalP analysis confirmed the presence of a signal peptide sequence in 324 proteins hereafter referred to as the salivary gland secretome (Supplemental File 9, Supporting Information).

GO term mapping was conducted on the *A. pisum* salivary gland secretome using the Blast2GO suite. At least one GO term could be assigned to 168 of the 324 proteins of the salivary gland secretome (Supplemental File 9, Supporting Information). The functional component of the salivary gland secretome was determined by categorizing the GO terms by MF, BP and CC (Figure 3). The three most common MF categories were protein binding (22.5%), peptidase activity (14.3%) and transporter activity (10.2%) whereas the three most common BP categories were regulation of biological process (17.4%) carbohydrate metabolic process (17.4%) and signal transduction (15.5%). The multilevel CC analysis returned three categories, the endoplasmic reticulum (44%), protein complex (23%) and extra-cellular region 23%.

One-hundred twenty-four proteins (38%) in the salivary gland secretome were deemed of unknown function after BLAST searches against non redundant databases, annotation augmentation based on mapped GO terms in BLAST2Go and manual checking of all proteins annotated as a hypothetical protein. Salivary gland contigs with a high R-value were BLASTx searched against the consensus ACYPI proteins, which resulted in the identification of protein matches for 91 contigs with high

Table 3. Analysis of the Evolution of a Selection of Predicted Pea Aphid Salivary Gland Proteins

salivary gene/family	number of paralogs ^a	rate of evolution ^{b,c}	dN/dS ^{b,d}	fraction of sites ^e	mode of selection
M1 metalloproteases	13	1.52***	6.92*	0.07	Diversifying
M2 metalloproteases	5	1.30 ^{NS}	33.8***	0.14	Diversifying
Glucose dehydrogenases	8	1.22*	99***	0.11	Diversifying
Apolipophorins	3	1.46***	1.30**	0.97	Diversifying
Glutathione peroxidase	1	0.71*	1.31 ^{NS}	0.40	Relaxed
Armet (ARP)	1	1.47*	2.08 ^{NS}	0.05	Relaxed
PR1-like protein	1	1.41***	1.00 ^{NS}	0.99	Relaxed
Prolyl-4-hydroxylase	1	0.98 ^{NS}	1.10 ^{NS}	<0.01	None
Regucalcin	1	0.96 ^{NS}	38 ^{NS}	0.01	None

^a Number of pea aphid paralogs included in analysis. Some paralogs were necessarily excluded because of insufficient amino acid overlap to allow analysis.

^b NS: not significant; *: $P > 0.90$; **: $P > 0.95$; ***: $P > 0.99$. ^c Rate of evolution relative to the most closely related homologue from another insect species, as identified from phylome analysis. ^d dN/dS calculated using branch-site model⁸⁰ ^e Fraction of sites upon which dN/dS value is calculated⁸⁰

R-statistic value (Supplemental File 8, Supporting Information). In some cases a number of different contigs matched to the same protein. In total, 42 individual proteins present in the salivary gland secretome (Table 1) were deemed to be enriched in the salivary glands. Thirty-two proteins supported by abundant EST contigs are currently nonannotatable. Supplemental File 10 (Supporting Information) describes proteins identified by mass spectrometry with secretion signal but without matches to salivary gland EST contigs with high R-statistic values. Of these 65 directly identified proteins that were supported by multiple peptides and predicted to have a secretion signal, 36 were present in the transcript-supported data set and 8 were deemed nonannotatable. Table 2 summarizes the principal saliva associated or pathogen effector-like proteins identified from this multidisciplinary approach.

Phylome Analysis of Saliva-Associated Genes

We selected 9 annotatable salivary gland genes (and gene families) predicted by the transcript enrichment and proteome methods for phylogenetic analysis (Table 3). Proteins were chosen based on a number of criteria including abundance in salivary gland, demonstration in both the salivary gland proteome and EST catalogues, demonstration in saliva either with mass spectrometry^{27,28} or with substrate specific assays (e.g., see refs 40–42), presence of a signal peptide sequence (albeit not in all cases) and similarity to known effector proteins from phytopathogenic organisms. The amino acid and coding sequences of all paralogs and a representative selection of insect homologues were included in the phylogenetic analysis (results not shown). Pea aphid paralogs were identified for the M1 and M2 metalloproteases, the ubiquitin specific proteases, the glucose dehydrogenases, the prolyl-4-hydroxylases, and the regucalcin (Table 3). Of the genes with nonsalivary paralogs, only the M1 metalloprotease genes showed a significantly higher rate of evolution relative to these paralogs. The M1 metalloproteases also evolved faster relative to insect homologues (Table 3) and the phylogenetic tree of the M1 metalloproteases is given as Figure 4. The glutathione peroxidase, Armet, PR-1 like protein, and both apolipophorins also evolved at a faster rate relative to homologues from other insects. Within the glucose dehydrogenases 24 sites were identified as being under positive selection (Figure 5) relative to other insect homologues. Seven nonsynonymous substitution sites were identified as evolving significantly faster than synonymous substitutions in the same genes, an indication that they are potentially under positive selection. Interestingly, the

two glucose dehydrogenases identified in this study (ACYPI000986 and ACYPI000288) also evolved at a faster rate relative to homologues in other insects, but in contrast the glucose dehydrogenase (ACYPI000113; GMC oxidoreductase) identified in pea aphid saliva²⁸ has evolved at a significantly lower rate relative to other insect homologues. We found no evidence for positive selection among the predicted salivary glucose dehydrogenase genes (ACYPI000986 and ACYPI000288). Only the glucose dehydrogenases, the apolipophorins, and the M2 metalloproteases were under positive selection using a branch-site model. In contrast to the other two, the positive selection observed on the M2 metalloproteases occurred at a small number of sites, without resulting in a high rate of evolution across the entire gene.

DISCUSSION

In this study we have adopted a multidisciplinary approach in an attempt to characterize the salivary gland secretome of the pea aphid *A. pisum* and to identify potential aphid effector proteins. Two clear trends are evident from this study of the pea aphid salivary gland secretome: (1) a high proportion of salivary gland proteins have signal peptide sequences, and (2) a high proportion of the proteins associated with aphid saliva (~50%) are of unknown function. This percentage is substantially higher than the average of nonannotatable proteins across the aphid genome (c.30%⁴³), but this is probably due to the highly specific function of aphid saliva or the strong selection pressure on aphids to adapt to the wide plethora of dynamic plant defense mechanisms. Indeed, most of these nonannotatable sequences matched ESTs from other aphid species.

Some of the discrepancies between the salivary gland transcriptome and proteome may be due to the use of different asexual lineages of pea aphid in this study. Different pea aphid clones elicit variable plant wound responses and *R*-gene resistance is known to be clone-specific^{13,45,46} and perhaps linked to clone-specific salivary elicitors. An additional problem with combining transcriptomic and proteomic data sets relates to the well-characterized differences that exist in the expression levels between the two systems. Expression profiles (mRNA) are highly dynamic, and mRNA expression and protein levels do not always correlate.⁴⁴ Comparisons of the transcriptomes and proteomes of *Caenorhabditis elegans* and *D. melanogaster* have demonstrated that a higher degree of correlation exists between the proteomes of these distantly related organisms than for the transcriptome and proteome of the same organism.⁴⁵ Thus variation between

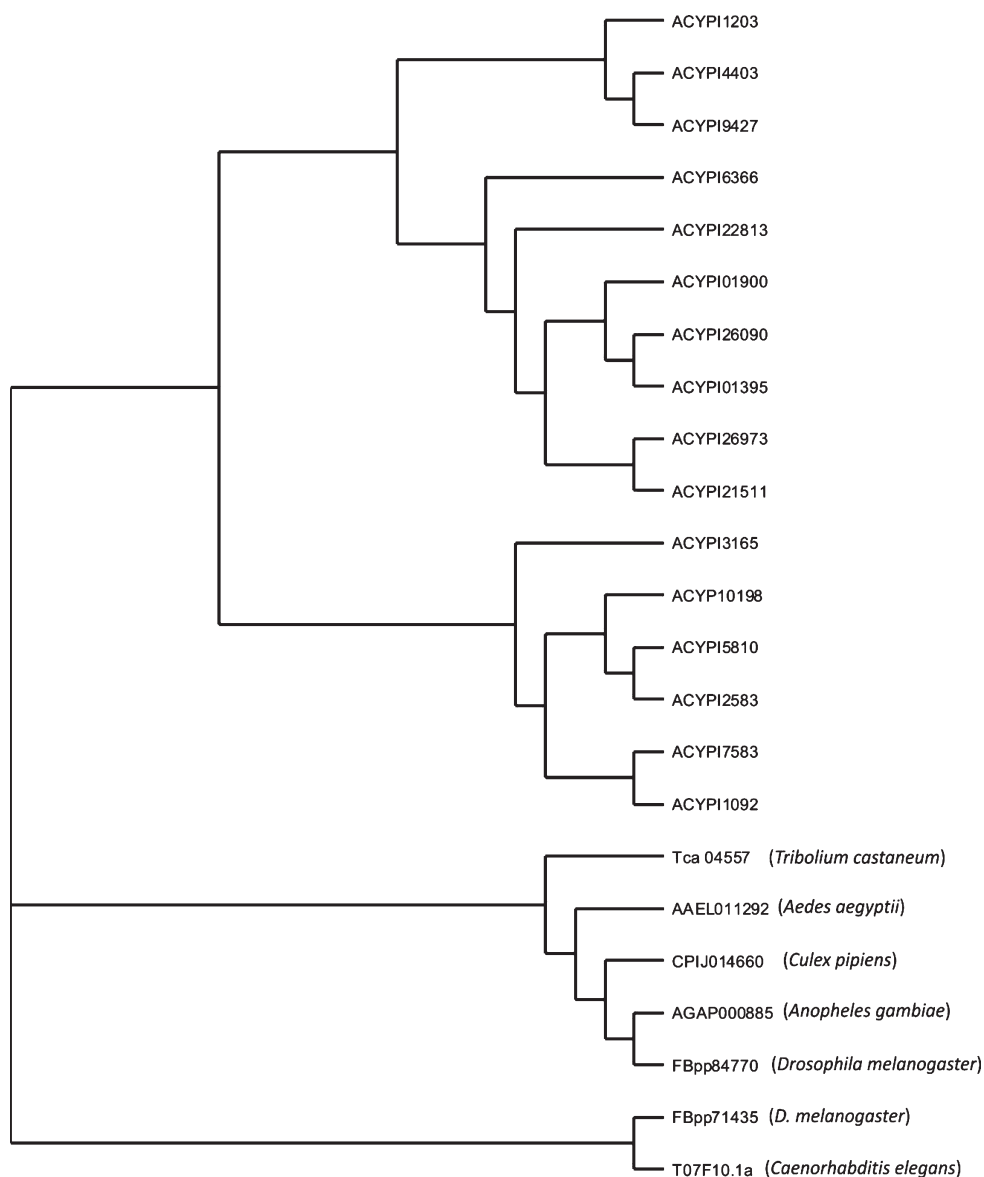


Figure 4. PhylomeDB analysis of predicted pea aphid salivary M1 metalloprotease ACYPI009427. The analysis identified 15 paralogs of this gene in the pea aphid genome, but only 7 homologues in other invertebrate genomes.

the expressed transcript and the translated protein will most likely be the norm rather than the exception. However the aim of our study was to report an initial catalogue of putative secreted salivary gland proteins based on transcriptomic and proteomic resources that are currently available. The combination of data sets reported here represents an important first step in identifying candidate effectors for future studies in aphids and other plant-sap feeding insects.

Our approach in predicting putative salivary gland proteins was based on two main assumptions. First, we assumed that transcripts for salivary proteins would be enriched in salivary gland EST libraries compared to whole organism libraries, and second, we assumed that salivary proteins would have an N-terminal secretion signal. The validity of this approach is supported by the fact that our resulting data set includes all but one (an unknown protein ACYPI008138) of the proteins identified by direct MS analysis of the saliva.²⁸ Further validation of our approach can be inferred from the high proportion of

annotatable salivary-gland enriched transcript supported proteins that possess a secretion signal. This includes some of the proteases, oxidoreductases, and apolipoprotein as well as some hypothetical proteins with homologues in other insect species. At the same time, several salivary-gland-enriched cDNAs are not predicted to be secreted including aquaporin, a eukaryotic translation factor subunit (EIF3), and a ubiquitin-specific protease. It is possible that salivary-gland enrichment of these transcripts could be associated with functions unrelated to secreted saliva. In addition, alternative mechanisms for the translocation of proteins into saliva may also exist and must also be considered. For example, not all proteins identified through MS analysis of collected saliva had secretion signals.²⁸ Consequently, it is likely that our assumptions have resulted in the exclusion of salivary proteins from our predicted pool of candidate effectors. For these reasons, we will consider some genes without secretion signals and some not supported by the R-statistics as part of this discussion.

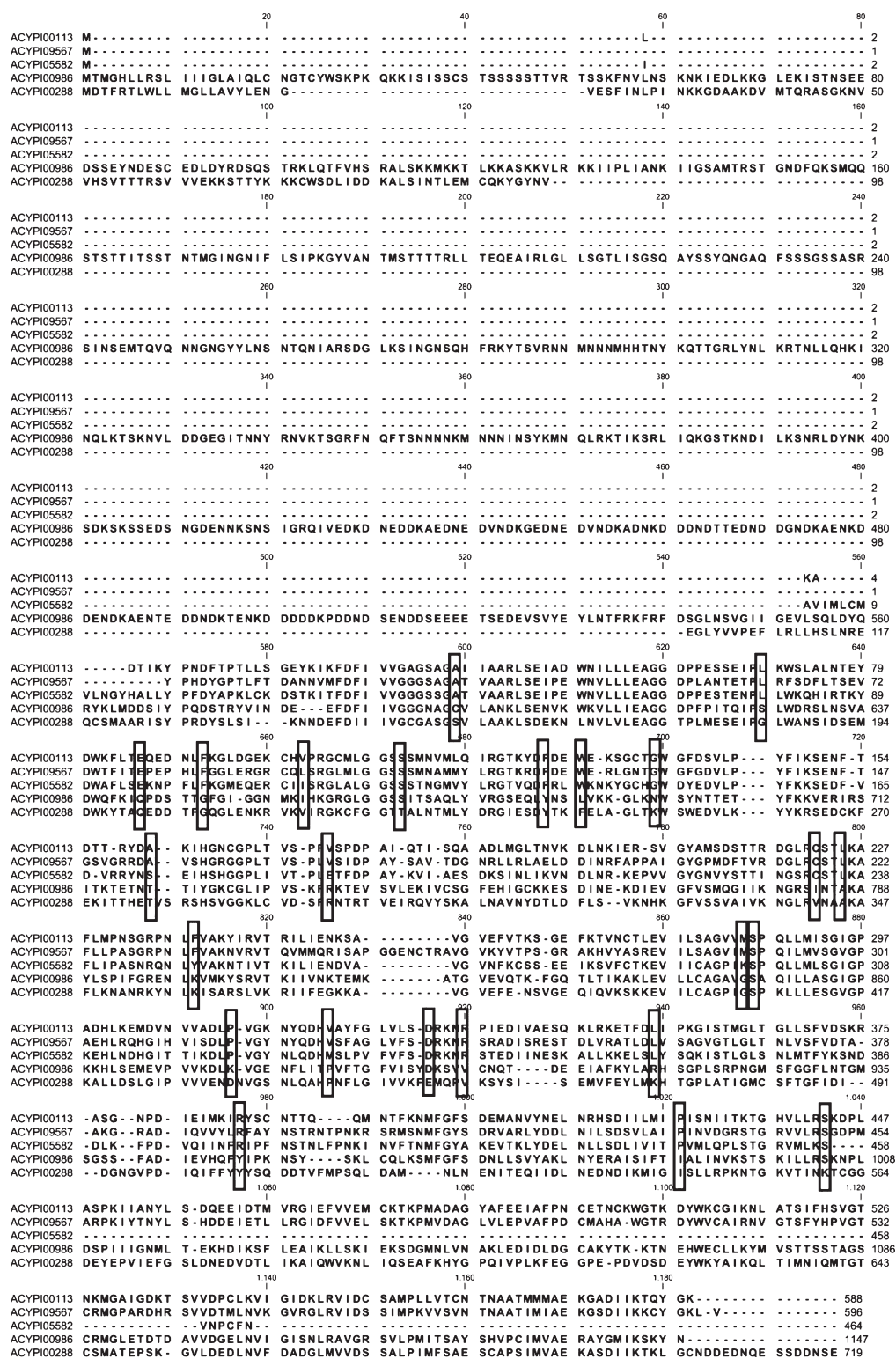


Figure 5. Amino acid alignment of five pea aphid glucose dehydrogenase paralogs identified from the pea aphid genome, including two paralogs (ACYPI000288, ACYPI000986) associated with salivary gland expression. Boxes indicate sites under positive selection (relative to homologues in other insects) a number of which represent non-synonymous substitutions specific to the predicted salivary gland paralogs.

The likelihood of false negatives would help to explain the small number of candidate salivary proteins we have identified in comparison with the number of secreted effectors identified from other eukaryotic pathogens. On the basis of their obligatory and specific association with host plants, aphids would be expected to

have a large number of effectors with subtle effects.⁴⁶ It is possible, however, that some salivary proteins could be produced in other organs. Hemolymph is constantly pumped in the salivary glands and it is likely that other biological molecules including proteins produced in other tissues make their way into the

salivary glands along with hemolymph⁴⁷ or by other, as yet unknown mechanisms. For example, no salivary gland transcripts or proteins identified in this study corresponded to the regucalcin protein reported previously in pea aphid saliva.²⁸ Searches for regucalcin EST transcripts in an *A. pisum* EST database (<http://www.aphidests.org/>) indicated distribution of regucalcin in the fat bodies and guts (results not shown). Additional research is necessary to confirm whether the salivary regucalcin is produced in the fat body, as is the case for homologous proteins in other insects.⁴⁸

Putative Effector Functions of Predicted Salivary Proteins

No functional data is yet available to provide insights into the role of any of the aphid salivary proteins predicted in this study. However, some potential effector roles for many of these candidate proteins can be predicted based on their homology or similarity to pathogenesis- or parasitism-related effector proteins secreted by other plant pathogens, and in particular plant-pathogenic nematodes. For the purpose of comparison, we will discuss our candidates in the context of the 8 categories proposed by Bellafiore et al.²⁴ for the secretome of the nematode *Meloidogyne incognita*. We found no candidate genes associated with cell wall modification or giant cell formation, which could be explained by the different feeding strategy of the pea aphid compared to plant parasitic nematodes. However, pectinase activity has been detected in the saliva of some aphids,^{41,42} and some aphids are thought to produce “pseudogalls” by altering host source-sink relationships.⁴⁹ It is possible these latter aphids could have proteins with homologous or analogous functions to the giant cell modifying proteins of plant parasitic nematodes.

Protein Synthesis and Secretion

We identified one protein with a possible protein folding function, the arginine-rich protein Armet, on our list of highly supported candidate salivary proteins. Armet (also called MANF) has been studied most in mammals, where it has both intracellular and extracellular functions.⁵⁰ Intracellularly, it is localized to the endoplasmic reticulum, where it appears to function as a component of the Unfolded Protein Response.⁵¹ It occurs in many tissues and organs but appears to occur at highest levels in proteins with intense protein secretion activities.⁵² It has previously been found to be enriched in salivary glands of sand flies (Supporting Information in ref S3). Armet is known to be a secreted protein,⁵⁴ and we have evidence that Armet is secreted into the plant during aphid feeding (Feng Cui, unpublished results). Several secreted protein disulfide isomerases (PDIs) and a cyclophilin were predicted by the ESTs and the proteome but were not found to be enriched in the salivary cDNA library (Table 1). PDIs and cyclophilins are both thought to be involved in regulating protein folding.^{55,56} They are also both present in the salivary secretions of plant parasitic nematodes, and PDIs have been associated with an increased yield of salivary proteins.⁵⁷

Detoxification

We also identified several candidate proteins that may be involved in reprogramming the plant cellular defense response. Reactive oxygen species (ROS) are produced as part of the initial defense response, and it has been proposed that oxidative enzymes in aphid saliva may act to degrade these ROS to help maintain a redox balance.²¹ We identified two glutathione peroxidases, key scavengers of ROS, among the enriched salivary gland transcripts. Two predicted glucose dehydrogenases (GLD,

GMC-oxidoreductase family) were present in our enriched transcript library, the salivary gland proteome, and were also identified directly in aphid saliva.²⁸ Insect GLDs have never been implicated in altering redox balance or otherwise modifying plant defenses, but other members of the insect GMC-oxidoreductase family have.^{58,59} A peroxinectin gene was also predicted from our analysis of salivary gland ESTs and proteome but was then not found to be significantly enriched in the salivary gland cDNA libraries. In invertebrates, peroxinectin is known to manipulate ROS during phagocytosis by regulating superoxide dismutase activity.⁶⁰ A protein similar to a plant pathogenesis-related protein (PR1) was also identified among our predicted secreted salivary proteins. PR1 proteins are thought to be lipid transfer proteins that mediate the signaling of systemic defense responses in plants.⁶¹ Secretion of insect PR1 homologues could interfere with this response. We also identified two lipid-binding apolipoproteins in our predicted secreted salivary proteins. Similar to PR1s, Apolipoproteins, after binding to lipid elicitor molecules, undergo a conformational change that results in an induction of the insect innate immune response.⁶² Again, it is conceivable that the secreted apolipoproteins could interfere with signaling of the plant's own cellular immune response.

Protein Modifications and Turnover

It has been proposed that plant-parasitic nematodes may selectively regulate protein degradation as part of their parasitism strategy.²⁶ We have identified a number of candidate pea aphid salivary proteins that may also function to effect host protein modification or turnover. M1 and M2 zinc metalloproteases were predicted by both the transcript and proteome analyses in this study, and were also identified at high abundance in an analysis of aphid saliva.²⁸ Metalloproteases are also common in the saliva of plant parasitic nematodes.²⁴ Both the M1 and M2 are predicted nonspecific proteases targeting the N terminus and C terminus, respectively, of short target peptides. Insect M2 metalloproteases are dipeptidyl peptidases that often target signaling peptides, such as hormones and neuropeptides.^{63,64} Both proteases may function to destroy plant defense proteins. Further research is needed to determine whether both salivary proteases retain a nonspecific protease activity or whether they may have been adapted to target particular defense-related proteins in plants. We also found a ubiquitin-specific protease (USP) among our enriched salivary gland ESTs. USPs are important contributors to protein turnover in cells, and can also function in the regulation of apoptosis.⁶⁵ Ubiquitin-associated proteins are often associated with pathogen secretomes, including nematodes.²⁴

Aphid-Specific Functions

A calcium-binding function has often been proposed for aphid salivary proteins.²³ Direct MS analysis of pea aphid salivary proteins identified a candidate calcium-binding protein, regucalcin,²⁸ which we have not identified in our analyses of the salivary gland. Four additional secreted calcium-binding proteins were predicted by our salivary gland ESTs and proteome (an endoplasmic, calreticulain and two calnexins), but were then not found to be enriched in the salivary gland cDNA libraries (Table S6, Supporting Information). Among the nonannotatable proteins in our predicted salivary set are some proteins identified in previous studies. One of these proteins, C002 (ACYPI008617), is known to be secreted by the pea aphid during feeding and is necessary for successful feeding on plants.^{29,33} The putative salivary sheath protein SHP²⁸ was also identified in both the enriched EST library and the salivary gland proteome. Two

abundant nonannotatable proteins, ACYPI008224 and ACYPI006346, have previously been identified in pea aphid saliva²⁸ signifying their potential importance in the saliva mediated interaction between plant and aphid. Two of our predicted nonannotatable salivary proteins (ACYPI56502 and ACYPI45001) are two of 13 paralogs identified in the pea aphid genome by PhylomeDB. These paralogs are all predicted to be secreted and are highly conserved but contain no known functional domains.

Evolution of Predicted Aphid Salivary Proteins

Ma and Guttman⁶⁶ have summarized the differences in effector evolution between eukaryotic and prokaryotic pathogens. Eukaryotic pathogen effector evolution is usually driven by diversifying selection and the generation and maintenance of multiple gene copies/alleles. In contrast, positive selection with rapid gene acquisition/gene loss is more often the driving force behind the evolution of prokaryotic pathogen effectors. Evolution of effectors is also usually faster in prokaryotes because of short generation times and horizontal gene transfer, whereas eukaryotic pathogens must rely on recombination as the predominant diversifying mechanism. Polymorphism mechanisms in eukaryotic effectors often result in truncated, nonfunctional alleles^{67,68} that can revert in response to future changes in selection pressure.⁶⁶ Aphids are eukaryotes, of course, but for most of their life cycle they reproduce asexually without recombination.⁶⁹ For many aphids, landscapes have been shown to be dominated by single “superclones”, providing little allelic diversity.⁷⁰ As such, aphids are an interesting test of the dichotomy in effector evolution between prokaryotic and eukaryotic pathogens. One might hypothesize that to make up for a lack of recombination aphids should have evolved mechanisms to accelerate effector evolution and/or to generate and maintain effector diversity.

Many of the annotatable salivary effectors predicted by our approach are evolving at a faster rate than their homologues in other insects. For those genes that exist as a single copy in the saliva, there is no evidence for positive selection suggesting that accelerated evolution is occurring by relaxation of purifying selection. It is possible these proteins have been recruited for a novel function in the saliva, and their original function may no longer be needed in aphids. In contrast, there is strong evidence of positive selection for all candidate effectors that have been associated with multiple gene copies in the pea aphid genome. This probably indicates that the duplicate copies are undergoing diversifying selection. For most of these duplicated effector genes positive selection is acting at a small number of sites either localized or distributed across the protein. For the M1 metalloproteases, the positive selection is acting in a short region of c. 70 amino acids spanning the predicted N-terminal start of the Peptidase M1 functional domain (results not shown). Positive selection was also restricted to a few sites in a c. 70 amino acid region of the M2 metalloproteases, within the predicted active site of the functional domain. The 7 nonsynonymous sites identified as being under positive selection in the glucose dehydrogenases were distributed within and between functional domains (Figure 5).

In contrast, the sites under positive selection in the apolipophorins are numerous and distributed quite evenly across the protein sequence, common to the two candidate copies, that was included in the analysis. The reason for this widespread positive selection on these proteins is not clear. Without information on the structures of these proteins, we are unable to determine whether these sites under positive selection are located in surface

regions that could be interacting directly with resistance proteins in their host-plant.⁴⁶

For all these candidate effectors associated with multiple gene copies, several additional paralogs also exist in the genome with no evidence yet that links them with the saliva. Many of the paralogs are truncated versions of the candidate salivary genes. Thus, generating and maintaining multiple copies of salivary effector genes appears to be one approach that aphids use to combat host defenses. Interestingly, some of these gene copies do not appear to be found in clusters within the genome; most copies are distributed among various genome scaffolds (data not shown). One driving force for gene copy dispersion is to escape both the adaptive and maladaptive effects of recombination.^{71–73} Amplified esterase genes in the green peach aphid are also dispersed throughout the genome, and are regulated epigenetically by DNA methylation.⁷⁴ Further research is necessary to determine whether duplicated copies of effector genes may be dispersed to escape epigenetic coregulation.

CONCLUSIONS

The successful colonization of plants by invading pathogens is determined by a range of effector proteins secreted into the host that suppress or modulate specific innate defenses.^{2,75} The aim of the study reported here was to determine if aphids have an analogous suite of potential effector proteins and if there is convergence between phloem-feeding and traditionally viewed phytophagous pathogens such as bacteria, viruses, nematodes, fungi and oomycetes. Four major characteristics are evident from the pea aphid salivary gland secretome: (1) a considerable proportion of salivary gland proteins have a peptide secretion signal; (2) c. 40% of the proteins in the salivary gland secretome are of unknown function or found only in aphids; (3) all but one directly identified salivary protein²⁸ are present within the salivary gland secretome; and (4) parallels exist between the composition of the aphid salivary secretome and the secretomes of phytopathogenic nematodes. Perhaps this latter characteristic is the most intriguing as it seems that a high degree of adaptive convergence has occurred within phylogenetically distant groups of plant pathogenic organisms. Encountering the same suite of innate defenses within the plant by different organisms has resulted in the employment of similar endogenous proteins, and many defining features of effector genes (e.g., members of gene families, fast rates of evolution etc.) are also present in the aphid salivary secretome. However, the high proportion of unique and novel proteins within the salivary secretome indicates that the specialized habit of phloem feeding has required the evolution of novel proteins with an as yet undetermined function.

Although aphids are not widely recognized as plant pathogens our initial characterization of the salivary gland secretome generates a list of potential effector proteins for future studies. Bacterial and fungal effectors have been well-characterized on a molecular level but very little is known about the identity and function of proteins secreted by phloem-feeding insects. The classification of plant pathogens is gradually expanding to include phloem-feeding insects such as aphids,⁷⁵ and central to this shift in view is the recognition of the molecular nature of the plant-aphid interaction and our increasing understanding of the composition of aphid saliva. The availability of these candidate effectors and saliva-associated proteins presents testable hypotheses that will lead to the determination of the function of aphid salivary proteins.

■ ASSOCIATED CONTENT

■ Supporting Information

Supplemental File 1, Supporting Information and methods. Supplemental File 2, EST read and assembly characteristics. Supplemental File 3, MS-identified peptides and proteins. Supplemental File 4, EST Blast2Go results. Supplemental File 5, Blast2Go results for transcript supported proteins. Supplemental File 6, Blast2Go results for MS-identified proteins. Supplemental File 7, gene ontology (GO) terms obtained the transcript-supported and MS-identified proteins. Supplemental File 8, High R-statistic valued salivary gland ESTs and their Blast hits. Supplemental File 9, pea aphid salivary gland secretome. Supplemental File 10, salivary gland secretome proteins identified by mass spectrometry, without matches to salivary gland EST contigs with high R-statistic values. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*James C. Carolan, Department of Biology, Callan Building, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland. Tel: +353-1-708 6845. Fax: +353-1-708 3845. E-mail: james.c.carolan@gmail.com.

Present Addresses

*Department of Biology, Callan Building, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland.

*Edenspace Systems Corporation, 1500 Hayes Drive, Manhattan, Kansas 66502, United States.

■ ACKNOWLEDGMENT

Supported by grants from the Arthropod Genomics Center (to G.R.R. and C. Culbertson), USDA CREES-NRI (grant G001-35302-00975 to G.R.R. and J.R.R.), the Kansas Agricultural Experiment Station (publication number 11-146-J), Science Foundation Ireland (grant 03/IN3/B381 to T.L.W.), Department of Agriculture Fisheries and Food (RSF07533 to T.L.W.) and an Irish Research Council for Science, Engineering and Technology PhD award to K.R. We thank A. E. Douglas for the provision of the pea aphid image and two anonymous reviewers for their many helpful comments on the manuscript. Access to and use of instrumentation of the UCD Conway Institute Mass Spectrometry Resource is gratefully acknowledged.

■ REFERENCES

- (1) Kessler, A.; Baldwin, I. T. Plant responses to insect herbivory: The emerging molecular analysis. *Annu. Rev. Plant Biol.* **2002**, *53* (1), 299–328.
- (2) Howe, G. A.; Jander, G. Plant Immunity to Insect Herbivores. *Annu. Rev. Plant Biol.* **2008**, *59* (1), 41–66.
- (3) Kaloshian, I.; Walling, L. L. Hemipterans as plant pathogens. *Annu. Rev. Plant Biol.* **2005**, *43* (1), 491–521.
- (4) Thompson, G. A.; Goggin, F. L. Transcriptomics and functional genomics of plant defence induction by phloem-feeding insects. *J. Exp. Bot.* **2006**, *57* (4), 755–766.
- (5) Jones, J. D. G.; Dangl, J. L. The plant immune system. *Nature* **2006**, *444* (7117), 323–329.
- (6) Heidel, A. J.; Baldwin, I. T. Microarray analysis of salicylic acid- and jasmonic acid-signalling in responses of *Nicotiana attenuata* to attack

by insects from multiple feeding guilds. *Plant Cell Environ.* **2004**, *27*, 1362.

(7) Tagu, D.; Klingler, J. P.; Moya, A. S.; Simon, J.-C. Early progress in aphid genomics and consequences for plant-aphid interactions studies. *Mol. Plant-Microbe Interact.* **2008**, *21* (6), 701–708.

(8) Rossi, M.; Goggin, F. L.; Milligan, S. B.; Kaloshian, I.; Ullman, D. E.; Williamson, V. M. The nematode resistance gene Mi of tomato confers resistance against the potato aphid. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9750.

(9) Dogimont, C.; Chovelon, V.; Tual, S.; Boissot, N.; Rittener, V.; Giovinnaz, N.; Bendahmane, A. Molecular diversity at the Vat/Pm-W resistance locus in melon. *Cucurbitaceae 2008, Proceedings of the IXth EUCARPIA meeting on genetics and breeding of Cucurbitaceae*; Pitrat, M, Ed.; INRA, Avignon, France, May 21–24th, **2008**.

(10) Milligan, S. B.; Bodeau, J.; Yaghoobi, J.; Kaloshian, I.; Zabel, P.; Williamson, V. M. The root knot nematode resistance gene Mi from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. *Plant Cell* **1998**, *10* (8), 1307–1320.

(11) Klingler, J.; Creasy, R.; Gao, L.; Nair, R. M.; Calix, A. S.; Jacob, H. S.; Edwards, O. R.; Singh, K. B. Aphid resistance in *Medicago truncatula* involves antixenosis and phloem-specific, inducible antibiosis, and maps to a single locus flanked by NBS-LRR resistance gene analogs. *Plant Physiol.* **2005**, *137* (4), 1445–1455.

(12) Klingler, J. P.; Edwards, O. R.; Singh, K. B. Independent action and contrasting phenotypes of resistance genes against spotted alfalfa aphid and bluegreen aphid in *Medicago truncatula*. *New Phytol.* **2007**, *173* (3), 630–640.

(13) Stewart, S. A.; Hodge, S.; Ismail, N.; Mansfield, J. W.; Feys, B. J.; Prosperi, J.-M.; Huguet, T.; Ben, C.; Gentzbittel, L.; Powell, G. The RAP1 gene confers effective, race-specific resistance to the pea aphid in *Medicago truncatula* independent of the hypersensitive reaction. *Mol. Plant-Microbe Interact.* **2009**, *22* (12), 1645–1655.

(14) Chisholm, S. T.; Coaker, G.; Day, B.; Staskawicz, B. J. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* **2006**, *124*, 803–814.

(15) Abramovitch, R. B.; Anderson, J. C.; Martin, G. B. Bacterial elicitation and evasion of plant innate immunity. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 601–611.

(16) McCann, H. C.; Guttman, D. S. Evolution of the type III secretion system and its effectors in plant-microbe interactions. *New Phytol.* **2008**, *177* (1), 33–47.

(17) Block, A.; Li, G.; Fu, Z. Q.; Alfano, J. R. Phytopathogen type III effector weaponry and their plant targets. *Curr. Opin. Plant Biol.* **2008**, *11*, 396–403.

(18) Birch, P. R.; Rehmany, A. P.; Pritchard, L.; Kamoun, S.; Beynon, J. L. Trafficking arms: Oomycete effectors enter host plant cells. *Trends Microbiol.* **2006**, *14*, 8–11.

(19) Rehmany, A. P.; Gordon, A.; Rose, L. E.; Allen, R. L.; Armstrong, M. R.; Whisson, S. C.; Kamoun, S.; Tyler, B. M.; Birch, P. R.; Beynon, J. L. Differential recognition of highly divergent downy mildew avirulence gene alleles by RPP1 resistance genes from two *Arabidopsis* lines. *Plant Cell* **2005**, *17* (6), 1839–1850.

(20) Dou, D.; Kale, S. D.; Wang, X.; Jiang, R. H.; Bruce, N. A.; Arredondo, F. D.; Zhang, X.; Tyler, B. M. RXLR-mediated entry of *Phytophthora sojae* effector Avr1b into soybean cells does not require pathogen-encoded machinery. *Plant Cell* **2008**, *20* (7), 1930–1947.

(21) Miles, P. W. Aphid saliva. *Biol. Rev.* **1999**, *74* (01), 41–85.

(22) Tjallingii, W. F. Salivary secretions by aphids interacting with proteins of phloem wound responses. *J. Exp. Bot.* **2006**, *57* (4), 739–745.

(23) Will, T.; Tjallingii, W. F.; Thonnessen, A.; van Bel, A. J. E. Molecular sabotage of plant defense by aphid saliva. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (25), 10536–10541.

(24) Bellafiore, S.; Shen, Z.; Rosso, M.-N.; Abad, P.; Shih, P.; Briggs, S. P. Direct Identification of the *Meloidogyne incognita* Secretome Reveals Proteins with Host Cell Reprogramming Potential. *PLoS Pathog.* **2008**, *4* (10), e1000192.

(25) Hussey, R. S. Disease-inducing secretions of plant-parasitic nematodes. *Annu. Rev. Phytopathol.* **1989**, *27* (1), 123–141.

- (26) Davis, E. L.; Hussey, R. S.; Baum, T. J. Getting to the roots of parasitism by nematodes. *Trends Parasit.* **2004**, *20*, 134–141.
- (27) Harmel, N.; Létocart, E.; Cherqui, A.; Giordanengo, P.; Mazzucchelli, G.; Guillonnet, F.; Pauw, E. D.; Haubruge, E.; Francis, F. Identification of aphid salivary proteins: a proteomic investigation of *Myzus persicae*. *Insect Mol. Biol.* **2008**, *17* (2), 165–174.
- (28) Carolan, J. C.; Fitzroy, C. I. J.; Ashton, P. D.; Douglas, A. E.; Wilkinson, T. L. The secreted salivary proteome of the pea aphid *Acyrtosiphon pisum* characterised by mass spectrometry. *Proteomics* **2009**, *9* (9), 2457–2467.
- (29) Mutti, N.; Park, Y.; Reese, J.; Reeck, G. RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrtosiphon pisum*. *J. Ins. Sci.* **2006**, *6*, 38.
- (30) Ramsey, J.; Wilson, A.; de Vos, M.; Sun, Q.; Tamborindeguy, C.; Winfield, A.; Malloch, G.; Smith, D.; Fenton, B.; Gray, S.; Jander, G. Genomic resources for *Myzus persicae*: EST sequencing, SNP identification, and microarray design. *BMC Genomics* **2007**, *8* (1), 423.
- (31) De Vos, M.; Jander, G. *Myzus persicae* (green peach aphid) salivary components induce defence responses in *Arabidopsis thaliana*. *Plant Cell Environ.* **2009**, *32* (11), 1548–1560.
- (32) Bos, J. I. B.; Prince, D.; Pitino, M.; Maffei, M. E.; Win, J.; Hogenhout, S. A. A functional genomics approach identifies candidate effectors from the aphid species *Myzus persicae* (Green Peach Aphid). *PLoS Genet.* **2010**, *6* (11), e1001216.
- (33) Mutti, N. S.; Louis, J.; Pappan, L. K.; Pappan, K.; Begum, K.; Chen, M.-S.; Park, Y.; Dittmer, N.; Marshall, J.; Reese, J. C.; Reeck, G. R. A protein from the salivary glands of the pea aphid, *Acyrtosiphon pisum*, is essential in feeding on a host plant. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (29), 9965–9969.
- (34) Rahbé, Y.; Febvay, G. Protein toxicity to aphids: an in vitro test on *Acyrtosiphon pisum*. *Entomol. Exp. Appl.* **1993**, *67*, 149–160.
- (35) Masoudi-Nejad, A.; Tonomura, K.; Kawashima, S.; Moriya, Y.; Suzuki, M.; Itoh, M.; Kanehisa, M.; Endo, T.; Goto, S. EGASebler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. *Nucleic Acids Res.* **2006**, *34*, W459–462.
- (36) Conesa, A.; Gotz, S.; Garcia-Gomez, J. M.; Terol, J.; Talon, M.; Robles, M. Blast2GO: a universal tool for annotation and visualization in functional genomics research. *Bioinformatics* **2005**, bti610.
- (37) Altschul, S.; Gish, W.; Miller, W.; Myers, E.; Lipman, D. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.
- (38) Stekel, D.; Git, Y.; Falciani, F. The comparison of gene expression from multiple cDNA libraries. *Genome Res.* **2000**, *10*, 2055–2061.
- (39) Thompson, J.; Gibson, T.; Plewniak, F.; Jeanmougin, F.; Higgins, D. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **1997**, *24*, 4876–4882.
- (40) Madhusudhan, V. V.; Miles, P. W. Mobility of salivary components as a possible reason for differences in the responses of alfalfa to the spotted alfalfa aphid and pea aphid. *Entomol. Exp. Appl.* **1998**, *86* (1), 25–39.
- (41) Cherqui, A.; Tjallingii, W. F. Salivary proteins of aphids; a pilot study on identification, separation and immunolocalization. *J. Insect Physiol.* **2000**, *46*, 1177–1186.
- (42) Ma, R.; Reese, J. C.; Black, W. C. I.; Bramel-Cox, P. Detection of pectinesterase and polygalacturonase from salivary secretions of living greenbugs, *Schizaphis graminum* (Homoptera: Aphididae). *J. Insect Physiol.* **1990**, *36*, 507–512.
- (43) The International Aphid Genomics. Genome Sequence of the Pea Aphid *Acyrtosiphon pisum*. *PLoS Biol.* **2010**, *8* (2), e1000313.
- (44) Nie, L.; Wu, G.; Zhang, W. Correlation between mRNA and protein abundance in *Desulfovibrio vulgaris*: A multiple regression to identify sources of variations. *Biochem. Biophys. Res. Commun.* **2006**, *339* (2), 603–610.
- (45) Schrimpf, S. P.; Weiss, M.; Reiter, L.; Ahrens, C. H.; Jovanovic, M.; Malmström, J.; Brunner, E.; Mohanty, S.; Lercher, M. J.; Hunziker, P. E.; Aebersold, R.; von Mering, C.; Hengartner, M. O. Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. *PLoS Biol.* **2009**, *7* (3), e1000048.
- (46) Dodds, P. N.; Lawrence, G. J.; Catanzariti, A. M.; Ayliffe, M. A.; Ellis, J. G. The *Melampsoralini* AvrLS67 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell* **2004**, *16*, 755–768.
- (47) Ponsen, M. B. Aphids. Their Biology, Natural Enemies and Control. In *World Crop Pests*; Minks, A. K., Harrewijn, P., Eds.; Elsevier: New York, 1987; Vol. 2A.
- (48) Nakajima, Y.; Natori, S. Identification and characterization of an anterior fat body protein in an insect. *J. Biochem.* **2000**, *127* (5), 901–908.
- (49) Burd, J. D.; Webster, J. A.; Puterka, G. J.; Hoxie, R. P.; Wellso, S. G. Effect of Russian wheat aphid on constituent and nonconstituent carbohydrate content in wheat seedlings. *Southwest. Entomol.* **1996**, *21*, 167–172.
- (50) Parkash, V.; Lindholm, P.; Peranen, J.; Kalkkinen, N.; Oksanen, E.; Saarma, M.; Leppanen, V.-M.; Goldman, A. The structure of the conserved neurotrophic factors MANF and CDFN explains why they are bifunctional. *Protein Eng. Des. Sel.* **2009**, *22* (4), 233–241.
- (51) Apostolou, A.; Shen, Y.; Liang, Y.; Luo, J.; Fang, S. Armet, a UPR-upregulated protein, inhibits cell proliferation and ER stress-induced cell death. *Exp. Cell Res.* **2008**, *314* (13), 2454–2467.
- (52) Lindholm, P.; Peränen, J.; Andressoo, J.-O.; Kalkkinen, N.; Kokaia, Z.; Lindvall, O.; Timmusk, T.; Saarma, M. MANF is widely expressed in mammalian tissues and differently regulated after ischemic and epileptic insults in rodent brain. *Mol. Cell. Neurosci.* **2008**, *39* (3), 356–371.
- (53) Valenzuela, J. G.; Garfield, M.; Rowton, E. D.; Pham, V. M. Identification of the most abundant secreted proteins from the salivary glands of the sand fly *Lutzomyia longipalpis*, vector of *Leishmania chagasi*. *J. Exp. Biol.* **2004**, *207* (21), 3717–3729.
- (54) Mizobuchi, N.; Hoseki, J.; Kubota, H.; Toyokuni, S.; Nozaki, J.-i.; Naitoh, M.; Koizumi, A.; Nagata, K. ARMET is a soluble ER protein induced by the unfolded protein response via ERSE-II element. *Cell Struct. Funct.* **2007**, *32* (1), 41–50.
- (55) Darby, N. J.; Penka, E.; Vincentelli, R. The multi-domain structure of protein disulfide isomerase is essential for high catalytic efficiency. *J. Mol. Biol.* **1998**, *276*, 239–247.
- (56) Schonbrunner, E. R.; Mayer, S.; Tropschug, M.; Fischer, G.; Takahashi, N.; Schmid, F. X. Catalysis of protein folding by cyclophilins from different species. *J. Biol. Chem.* **1991**, *266* (6), 3630–3635.
- (57) Geldhof, P.; Vercauteren, I.; Knox, D.; Demare, V.; Van Zeveren, A.; Berx, G.; Vercruysse, J.; Claerebout, E. Protein disulphide isomerase of *Ostertagia ostertagi*: an excretory-secretory product of L4 and adult worms? *Int. J. Parasitol.* **2003**, *33* (2), 129–136.
- (58) Musser, R. O.; Hum-Musser, S. M.; Eichenseer, H.; Peiffer, M.; Ervin, G. Herbivory: Caterpillar saliva beats plant defences—a new weapon emerges in the evolutionary arms race between plants and herbivores. *Nature* **2002**, *416*, 599.
- (59) Eichenseer, H.; Mathews, M. C.; Bi, J. L.; Murphy, J. B.; Felton, G. W. Salivary glucose oxidase: multifunctional roles for *Helicoverpa zea*? *Arch. Insect Biochem. Physiol.* **1999**, *42*, 99–109.
- (60) Holmblad, T.; Söderhäll, K. Cell adhesion molecules and antioxidative enzymes in a crustacean, possible role in immunity. *Aquaculture* **1999**, *172* (1–2), 111–123.
- (61) Maldonado, A. M.; Doerner, P.; Dixon, R. A.; Lamb, C. J.; Cameron, R. K. A putative lipid transfer protein involved in systemic resistance signalling in *Arabidopsis*. *Nature* **2002**, *419* (6905), 399–403.
- (62) Whitten, M. M. A.; Tew, I. F.; Lee, B. L.; Ratcliffe, N. A. A novel role for an insect apolipoprotein (Apolipoprotein III) in {beta}-1,3-lucan pattern recognition and cellular encapsulation reactions. *J. Immunol.* **2004**, *172* (4), 2177–2185.
- (63) Lamango, N. S.; Nachman, R. J.; Hayes, T. K.; Strey, A.; Isaac, R. E. Hydrolysis of insect neuropeptides by an angiotensin-converting enzyme from the housefly, *Musca domestica*. *Peptides* **1997**, *18* (1), 47–52.
- (64) Isaac, R. E.; Schoofs, L.; Williams, T. A.; Veelaert, D.; Sajid, M.; Corvol, P.; Coates, D. A novel peptide-processing activity of insect

peptidyl-dipeptidase A (angiotensin I-converting enzyme): the hydrolysis of lysyl-arginine and arginyl-arginine from the C-terminus of an insect prohormone peptide. *Biochem. J.* **1998**, 330, 61–65.

(65) YANG, J.-m. Emerging roles of deubiquitinating enzymes in human cancer. *Acta Pharmacol. Sin.* **2007**, 28 (9), 1325–1330.

(66) Ma, W.; Guttman, D. S. Evolution of prokaryotic and eukaryotic virulence effectors. *Curr. Opin. Plant Biol.* **2008**, 11 (4), 412–419.

(67) Luderer, R.; Takken, F. L. W.; Wit, P. J. G. M. d.; Joosten, M. H. A. J. *Cladosporium fulvum* overcomes Cf-2-mediated resistance by producing truncated AVR2 elicitor proteins. *Mol. Microbiol.* **2002**, 45 (3), 875–884.

(68) Stergiopoulos, L.; De Kock, M. J. D.; Lindhout, P.; De Wit, P. J. G. M. Allelic variation in the effector genes of the tomato pathogen *Cladosporium fulvum* reveals different modes of adaptive evolution. *Mol. Plant-Microbe Interact.* **2007**, 20 (10), 1271–1283.

(69) Hales, D. F.; Tomiuk, J.; Woehrmann, K.; Sunnucks, P. Evolutionary and genetic aspects of aphid biology: A review. *Eur. J. Entomol.* **1997**, 94 (1), 1–55.

(70) Vorburger, C.; Lancaster, M.; Sunnucks, P. Environmentally related patterns of reproductive mode in the aphid *Myzus persicae* and the predominance of two superclones in Victoria, Australia. *Mol. Ecol.* **2003**, 12, 3493–3504.

(71) Katju, V.; Lynch, M. The structure and early evolution of recently arisen gene duplicates in the *Caenorhabditis elegans* Genome. *Genetics* **2003**, 165 (4), 1793–1803.

(72) Pal, C.; Hurst, L. D. Evidence for co-evolution of gene order and recombination rate. *Nat. Genet.* **2003**, 33 (3), 392–395.

(73) Taylor, J. S.; Raes, J. Duplication and Divergence: The evolution of new genes and old ideas. *Annu. Rev. Genet.* **2004**, 38 (1), 615–643.

(74) Field, L. M. Methylation and expression of amplified esterase genes in the aphid *Myzus persicae* (Sulzer). *Biochem. J.* **2000**, 349 (3), 863–868.

(75) Hogenhout, S. A.; Van der Hoorn, R. A. L.; Terauchi, R.; Kamoun, S. Emerging concepts in effector biology of plant-associated organisms. *Mol. Plant-Microbe Interact.* **2009**, 22 (2), 115–122.

(76) Jones, J. T.; Reavy, B.; Smant, G.; Prior, A. E. Glutathione peroxidases of the potato cyst nematode *Globodera rostochiensis*. *Gene* **2004**, 324, 47–54.

(77) Gotz, S.; Garcia-Gomez, J.; Terol, J.; Williams, T.; Nagaraj, S.; Nueda, M.; Robles, M.; Talon, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **2008**, 36 (10), 3420–3435.

(78) Chen, M.-S.; Zhao, H.-X.; Zhu, Y. C.; Scheffler, B.; Liu, X.; Liu, X.; Hulbert, S.; Stuart, J. J. Analysis of transcripts and proteins expressed in the salivary glands of Hessian fly (*Mayetiola destructor*) larvae. *J. Insect Physiol.* **2008**, 54 (1), 1–16.

(79) Brodmann, D.; Schuller, A.; Ludwig-Müller, J.; Aeschbacher, R. A.; Wiemken, A.; Boller, T.; Wingler, A. Induction of trehalase in *Arabidopsis* plants infected with the trehalose-producing pathogen *Plasmodiophora brassicae*. *Mol. Plant-Microbe Interact.* **2007**, 15 (7), 693–700.

(80) Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **2007**, 24 (8), 1586–1591.